DOI:10.11931/guihaia.gxzw202307014

蛇足石杉 *COBRA* 基因家族的分子生物信息学及表达分析

黄玉妹，滕建北，涂冬萍*，梁柳观

（广西中医药大学 药学院，南宁 530200）

摘　要：弄清蛇足石杉 *COBRA* 基因家族成员分子生物信息学特征及组织表达规律，为 *COBRA* 基因进一步研究提供依据。基于蛇足石杉的全长转录组数据，通过生物信息学技术对该家族成员（HsCOBRAs）的理化性质、结构域、保守基序、顺式作用元件、基因表达量等进行分析。结果表明：（1）在蛇足石杉全长转录组中共筛选出 24 个 HsCOBRAs 家族成员，其中酸性蛋白 9 个，稳定蛋白 11 个，疏水性蛋白 5 个，具有跨膜结构的蛋白 7 个，具有信号肽的蛋白 3 个。（2）亚细胞定位在细胞壁、叶绿体、细胞核、细胞膜上。（3）结构分析发现 HsCOBRAs 有 7 种结构域和 6 种保守基序，部分成员具有高度保守的 CCVS 结构。（4）HsCOBRAs 具有 CAAT-box、TATA-box 等 46 种顺式作用元件。（5）*HsCOBRA*2 在叶、孢子、茎、芽胞中的表达量均最高。蛇足石杉 *COBRA* 基因家族分子生物信息学及表达特征可为 *HsCOBRAs* 的进一步研究及生物学功能验证等提供理论依据。

关键词：蛇足石杉，*COBRA* 基因家族，生物信息学，全长转录组，表达分析

中图分类号：Q943　　　　　　　　**文献标识码**：A

Molecular bioinformatics and expression analysis of the *COBRA* gene family in *Huperzia serrata*

HUANG Yumei, TENG Jianbei, TU Dongping*, LIANG Liuguan

(*College of Pharmacy, Guangxi University of Traditional Chinese Medicine*, Nanning 530200, China)

Abstract: The paper clarifies the molecular bioinformatics characteristics and tissue expression patterns of the *COBRA* gene family members of *Huperzia serrata*, to provide a basis for further research on *COBRA* genes. Based on the full-length transcriptome data of the *Huperzia serrata*, the physicochemical properties, domains, conserved motifs, *cis*-acting elements, and genes expression of the family members (HsCOBRAs) were analyzed by bioinformatics techniques. The results were as follows: (1) A total of 24 HsCOBRAs family members were screened in the full-length transcriptome of *Huperzia serrata*, including 9 acidic proteins, 11 stabilizing proteins, 5 hydrophobic proteins, 7 proteins with transmembrane structures, and 3 proteins with signal peptides. (2) Subcellular localization was found in the cell wall, chloroplast, nucleus, and cell membrane. (3) Structural analysis revealed that HsCOBRAs have 7 domains and 6 conserved motifs, and partial members have a highly conserved CCVS structure. (4) HsCOBRAs have 46 *cis*-acting elements such as CAAT-box and TATA-box. (5) *HsCOBRA*2 had the highest expressionin in leaves, spores, stems and gemma. The molecular bioinformatics and expression

---

基金项目：广西自然科学基金项目（2022JJA140897）；广西中医药大学桂派中医药传承创新团队（2022B005）；广西中医药大学桂派杏林青年英才（2022C033）；广西中医药大学研究生教育创新计划项目（YCSY2022007）。

第一作者：黄玉妹（1998-），硕士研究生，研究方向为中药（壮瑶药）鉴定，（E-mail）1989543360@qq.com。

***通信作者**：涂冬萍，博士，副教授，研究方向为中药、民族药的品质研究，（E-mail）fairytdp@ 163.com。

characterization of the *COBRA* gene family of *Huperzia serrata* can provide theoretical basis for further research and biological function verification of *HsCOBRAs*.
Key words: *Huperzia serrata*, *COBRA* gene family, bioinformatics, full-length transcriptome, expression analysis

COBRA 基因最早发现于拟南芥根细胞中(Borner et al., 2003)，是一类广泛存在于高等植物和藻类植物(任昂彦等，2021)中的纤维素合成酶基因(Kasirajan et al., 2018)，其编码的蛋白属于糖基磷脂酰肌醇锚定蛋白(glycosyl-phosphatidyl inositol, GPI)，包含 N 端信号肽、碳水化合物结合域、富含半胱氨酸的 CCVS 结构域和 C 末端 GPI 蛋白的ω-位点(Roudier et al., 2002)。目前，学者们已鉴定了棉花、大白菜、茶树、番茄、玉米等多种植物的 *COBRA* 基因。研究表明 *COBRA* 可通过调节纤维素微纤丝的合成及沉积，在初生和次生细胞壁纤维素的生物合成中发挥核心作用(Li et al., 2019；任昂彦等，2021)，并广泛参与根、茎、叶、花及果实的发育，影响着植物的生物量(元志成等，2020；Zaheer et al., 2022; Li et al., 2022)。如棉花 *GhCOBL9A* 的过量表达可使生物量增加(Niu et al., 2018)；茶树 *CsCOBRA* 基因可通过调控叶片细胞壁及机械强度，从而影响茶叶的产量(艾安涛等，2021)。

蛇足石杉(*Huperzia serrata*) 是中药材千层塔的来源之一，具有活血行瘀、止血生肌、消肿定痛、清热解毒、退热除湿等功效，用于治疗跌打损伤、劳伤吐血、痔疮便血、肺痈、烧烫伤等（黄玉妹等，2023）。其主要成分石杉碱甲对阿尔茨海默病（alzheimer’s disease, AD）具有良好的疗效，且外周不良反应轻微(姬生国，2007)。预计到 2050 年，全球 AD 患病人数将达 1.39 亿(Jia et al., 2018)，我国老年人患病人数将达 3 003 万(王英全等，2019)，蛇足石杉需求量不断增大，市场前景良好。近年来，蛇足石杉的人工栽培取得一定进展，但因其生长缓慢，繁殖困难，至今尚未实现大规模栽培，产量远远不能满足市场需求。高产量、高效益的蛇足石杉现代化生产已成迫切需要，通过 *COBRA* 等纤维素合成酶基因的调控有望成为蛇足石杉生物产量增长的新途径。

本研究以蛇足石杉为研究材料，从其全长转录组测序数据中筛选 *COBRA* 基因，通过生物信息学技术对基因特征及在组织中的表达量进行分析，探讨：（1）蛇足石杉 *COBRA* 基因家族的生物信息学特征；（2）*COBRA* 基因在蛇足石杉生长发育中可能发挥的作用，为将来蛇足石杉 *COBRA* 基因功能的鉴定及调控奠定基础，为蛇足石杉的分子育种从而提高产量提供依据。

1 材料与方法

1.1 材料

蛇足石杉植株于 2019 年 8 月在广西南宁市大明山采集，经广西药用植物园冯世鑫高级工程师鉴定为石杉科植物蛇足石杉(*Huperzia serrata*) 的全草。蛇足石杉全长转录组数据由深圳华大基因股份有限公司 PacBio Sequel 平台提供。

1.2 方法

1.2.1 蛇足石杉 *COBRA* 基因家族成员鉴定

基于蛇足石杉的全长转录组测序数据及其注释结果，利用拟南芥 *COBRA* 基因家族成员的 DNA 序列作为种子序列，BLAST 比对搜索蛇足石杉全基因序列数据库，筛选阈值 $E \leqslant 10^{-5}$，获得蛇足石杉的 *COBRA* 基因家族成员。

1.2.2 蛇足石杉 *COBRA* 基因家族的理化性质、信号肽、亚细胞定位和结构域分析

将蛇足石杉 *COBRA* 基因家族的原始碱基序列通过开放阅读框 ORF Finder（http://www.ncbi.nlm.nih.gov/orffinder)获得潜在的蛋白质编码片段。利用 ExPAsy(Expasy - ProtParam tool）对 *COBRA* 基因编码的蛋白质序列进行理化性质预测，SignalP-6.0

（https://services.healthtech.dtu.dk/service.php?SignalP）进行信号肽预测，Plant-mPLoc（http://www.csbio.sjtu.edu.cn/bioinf/plant-multi/）进行亚细胞定位预测，Batch CD-Search（https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi）进行结构域预测，并利用TBtools1.112 软件进行可视化处理。

1.2.3 蛇足石杉 *COBRA* 基因家族的二级结构、三级结构和顺式作用元件分析

分别通过 Prabi（https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html）、phyre^2（http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index）进行二级结构和三级结构预测。以蛇足石杉 *COBRA* 家族基因上游 2 000 pb 序列作为启动子区域，利用 Plant CARE（https://bioinformatics.psb.ugent.be/webtools/plantcare/html/）进行顺式作用元件预测，筛选具有注释功能的顺式作用元件，利用 TBtools1.112 软件进行可视化处理。

1.2.4 蛇足石杉 *COBRA* 基因家族的保守基序、表达量和系统发育分析

通过 MEME Suite 5.5.1（https://meme-suite.org/meme/tools/meme）预测基因家族的保守基序，利用 TBtools1.112 软件进行可视化处理。通过计算每千个碱基的转录每百万映射读取的片段(fragments per kilobase of exon model per million mapped fragments，FPKM) 均值，分析 *COBRA* 基因家族在蛇足石杉叶、孢子、茎、芽胞中的表达量，利用 TBtools 1.112 软件绘制基因表达量热图并进行美化。通过 MEGA11 软件对蛋白序列进行多重比较，使用邻接法（neighbor-joining)构建系统进化树（bootstrap=1 000)，利用 iTOL（http://itol.embl.de）对进化树进行美化。

2 结果与分析

2.1 蛇足石杉 *COBRA* 基因家族的理化性质分析

从蛇足石杉全长转录组数据中共筛选出 24 条 *COBRA* 基因家族的碱基序列，分别命名为 *HsCOBRA*1~*HsCOBRA*24，其碱基序列可通过图 1：A 获得。基于蛇足石杉转录组测序及 ORFfinder 获得氨基酸序列（图 1：B）。利用 ExPAsy 在线软件对 *HsCOBRAs* 编码的蛋白质进行理化性质预测，结果见表 1。24 条 HsCOBRAs 蛋白序列中，蛋白氨基酸数量为 73（HsCOBRA19）~663（HsCOBRA14）个。分子量在 8 494.91（HsCOBRA15）~73 084.23（HsCOBRA14）Da 之间，平均分子量为 26 012.36。9 个蛋白（HsCOBRA1、HsCOBRA5、HsCOBRA6、HsCOBRA7、HsCOBRA9、HsCOBRA14、HsCOBRA16、HsCOBRA17、HsCOBRA18）等电点在 4.53~6.97 之间，为酸性蛋白；其余 15 个蛋白的等电点在 8.17~10.13 之间，为碱性蛋白。11 个蛋白（HsCOBRA1、HsCOBRA2、HsCOBRA3、HsCOBRA4、HsCOBRA6、HsCOBRA7、HsCOBRA11、HsCOBRA12、HsCOBRA13、HsCOBRA14、HsCOBRA23）为稳定蛋白，其余 13 个蛋白均为不稳定蛋白。HsCOBRAs 家族成员的脂溶性指数介于 61.70~92.37 之间；HsCOBRA7、HsCOBRA8、HsCOBRA9、HsCOBRA12、HsCOBRA15 为疏水性蛋白；其余蛋白为亲水性蛋白。所有蛋白均带有负电残基 3（HsCOBRA19）~77（HsCOBRA16）个及正电残基 6（HsCOBRA15）~53（HsCOBRA5）个。

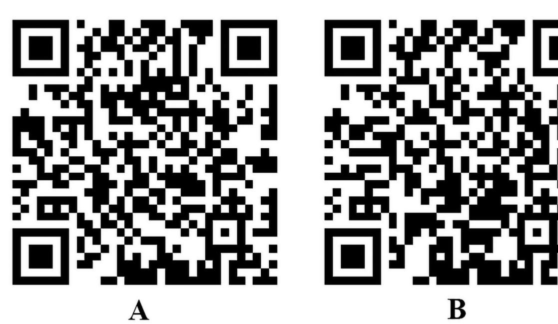


图 1 蛇足石杉 COBRA 家族的碱基（A）及氨基酸序列（B）二维码

Fig.1 The two-dimensional code of COBRA family base sequence（**A**）and amino acid sequence（**B**） of *Huperzia serrata*

表 1 蛇足石杉 *COBRA* 家族理化性质分析

Table 1 Physicochemical characteristics of *COBRA* family of *Huperzia serrata*

基因名 Gene name	氨基酸数量 Number of amino acids	分子量 Molecular weight (Da)	等电点 pI	不稳定系数 Instability index	亲脂系数 Aliphatic index	亲水性平均系数 GRAVY	带负电残基 (Asp + Glu)	带正电残基 (Arg + Lys)
*HsCOBRA*1	302	32 530.59	5.47	36.85	65.26	-0.420	37	28
*HsCOBRA*2	471	52 393.40	8.17	38.96	81.80	-0.042	28	32
*HsCOBRA*3	503	55 930.98	8.68	38.03	68.09	-0.246	32	42
*HsCOBRA*4	471	52 393.40	8.17	38.96	81.80	-0.042	28	32
*HsCOBRA*5	497	54 926.58	5.20	46.82	83.38	-0.208	66	53
*HsCOBRA*6	122	13 210.99	4.99	32.01	74.43	-0.009	15	10
*HsCOBRA*7	120	13 043.86	5.02	31.76	78.92	0.041	16	10
*HsCOBRA*8	142	15 656.18	8.78	53.72	90.00	0.008	7	11
*HsCOBRA*9	137	15 292.54	4.58	42.90	91.75	0.120	12	8
*HsCOBRA*10	81	9 655.12	9.03	42.43	69.88	-0.267	7	10
*HsCOBRA*11	255	28 653.78	8.24	39.64	68.47	-0.209	15	18
*HsCOBRA*12	154	16 979.79	8.35	38.91	79.81	0.102	8	11
*HsCOBRA*13	342	37 845.55	8.66	39.42	75.61	-0.080	17	24
*HsCOBRA*14	663	73 084.23	5.05	35.02	83.91	-0.076	57	43
*HsCOBRA*15	75	8 494.91	8.74	56.17	71.47	0.128	4	6
*HsCOBRA*16	80	9 176.60	6.52	53.67	80.50	-0.094	77	7
*HsCOBRA*17	127	13 913.37	4.53	47.61	71.34	-0.412	22	8
*HsCOBRA*18	174	19 703.91	6.97	46.39	84.08	-0.010	12	12
*HsCOBRA*19	73	8 531.23	10.13	61.70	80.27	-0.255	3	13
*HsCOBRA*20	141	15 644.73	9.57	65.95	61.70	-0.397	7	16
*HsCOBRA*21	119	13 578.72	9.52	42.76	65.55	-0.606	13	21
*HsCOBRA*22	232	25 716.16	8.47	51.62	73.10	-0.656	29	31
*HsCOBRA*23	194	22 823.72	9.23	18.98	92.37	-0.343	24	30
*HsCOBRA*24	130	15 116.39	8.95	61.45	85.38	-0.189	6	11

2.2 蛇足石杉 *COBRA* 基因家族亚细胞定位及信号肽分析

跨膜结构预测发现 7 个蛋白（HsCOBRA2、HsCOBRA3、HsCOBRA4、HsCOBRA9、HsCOBRA11、HsCOBRA12、HsCOBRA14）分别具有 1 个跨膜结构，其余均无跨膜结构。信号肽预测发现 HsCOBRA2、HsCOBRA4、HsCOBRA14 具有典型的 GPI 锚定蛋白 N 端信号肽，信号肽指数分别为 0.975 8、0.975 8、0.999 8，其切割位点分别可能在 N 端的第 41 和第 42、第 41 和第 42、第 28 和第 29 个氨基酸残基间；其余信号肽指数均小于 0.5。亚细胞定位发现 HsCOBRAs 家族成员分布广泛，1 个（HsCOBRA9）定位于细胞壁，4 个（HsCOBRA6、HsCOBRA7、HsCOBRA10、HsCOBRA23）定位于叶绿体，10 个（HsCOBRA1、

HsCOBRA5、HsCOBRA7、HsCOBRA16、HsCOBRA17、HsCOBRA19、HsCOBRA20、HsCOBRA21、HsCOBRA22、HsCOBRA24）定位于细胞核，其余 10 个均定位于细胞膜。HsCOBRA7 分别定位于叶绿体、细胞核，HsCOBRA10 分别定位于叶绿体、过氧化物酶体，详见表 2。

表 2 蛇足石杉 COBRA 家族亚细胞定位及信号肽分析

Table 2 Subcellular localization and signal peptide analysis of COBRA family of *Huperzia serrata*

蛋白 Protein	亚细胞定位 Subcellular location	跨膜结构 Transmembrane structure	信号肽 Signal peptide	蛋白 Protein	亚细胞定位 Subcellular location	跨膜结构 Transmembrane structure	信号肽 Signal peptide
HsCOBRA1	细胞核 Nucleus	0	0.000 8	HsCOBRA13	细胞膜 Cell membrane	0	0.002 5
HsCOBRA2	细胞膜 Cell membrane	1	0.975 8	HsCOBRA14	细胞膜 Cell membrane	1	0.999 8
HsCOBRA3	细胞膜 Cell membrane	1	0.000 3	HsCOBRA15	细胞膜 Cell membrane	0	0.001 0
HsCOBRA4	细胞膜 Cell membrane	1	0.975 8	HsCOBRA16	细胞核 Nucleus	0	0.000 1
HsCOBRA5	细胞核 Nucleus	0	0	HsCOBRA17	细胞核 Nucleus	0	0.000 1
HsCOBRA6	叶绿体 Chloroplast	0	0.000 8	HsCOBRA18	细胞膜 Cell membrane	0	0.000 4
HsCOBRA7	叶绿体，细胞核 Chloroplast，nucleus	0	0.000 8	HsCOBRA19	细胞核 Nucleus	0	0
HsCOBRA8	细胞膜 Cell membrane	0	0	HsCOBRA20	细胞核 Nucleus	0	0
HsCOBRA9	细胞壁 Cell wall	1	0	HsCOBRA21	细胞核 Nucleus	0	0
HsCOBRA10	叶绿体，过氧化物酶体 Chloroplast，peroxisome	0	0.075 1	HsCOBRA22	细胞核 Nucleus	0	0
HsCOBRA11	细胞膜 Cell membrane	1	0.000 1	HsCOBRA23	叶绿体 Chloroplast	0	0
HsCOBRA12	细胞膜 Cell membrane	1	0.000 1	HsCOBRA24	细胞核 Nucleus	0	0.000 3

2.3 蛇足石杉 *COBRA* 基因家族结构域分析

蛋白结构域分析结果显示 HsCOBRAs 家族共有 7 种结构域（图 2），分别为 COBRA、COBRA super family、PI31_Prot_N、PI31_Prot_C、PI31_Prot_N super family、RING_Ubox

super family、TilS super family。其中 COBRA 与 COBRA super family 分别分布在 4 个蛋白序列中，均注释为糖磷脂酰肌醇。除 HsCOBRA1 含有 2 个结构域外，13 个蛋白（HsCOBRA2、HsCOBRA3、HsCOBRA4、HsCOBRA5、HsCOBRA6、HsCOBRA7、HsCOBRA8、HsCOBRA12、HsCOBRA13、HsCOBRA14、HsCOBRA16、HsCOBRA17、HsCOBRA18）仅有一个结构域，其余的 10 个蛋白无结构域。

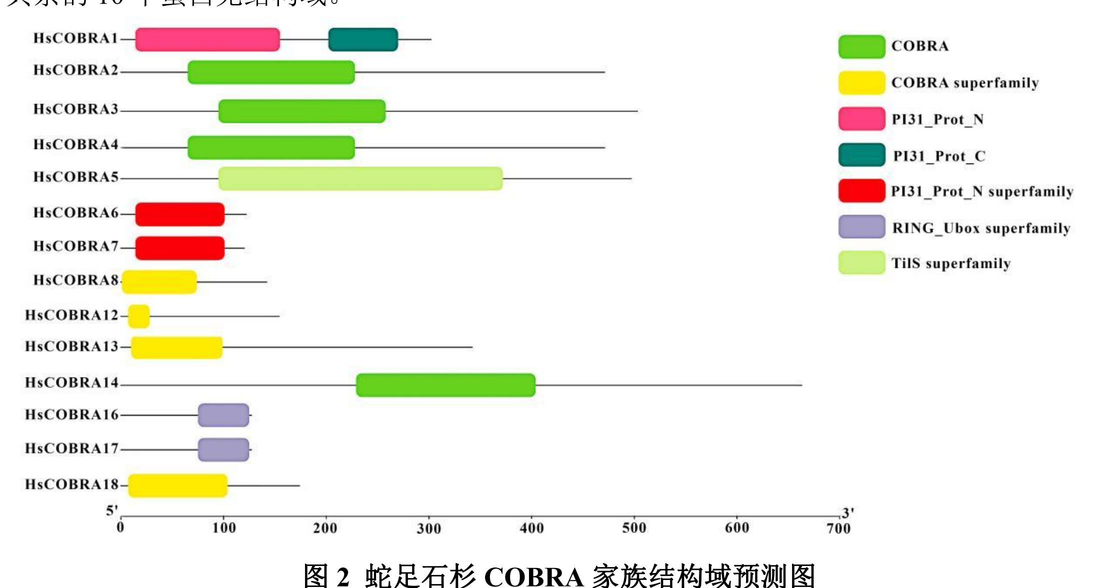


图 2 蛇足石杉 COBRA 家族结构域预测图

Fig. 2 The COBRA family domain prediction of *Huperzia serrata*

2.4 蛇足石杉 *COBRA* 基因家族二级结构分析

利用 Prabi 软件对蛇足石杉 *COBRA* 基因家族进行二级结构预测，表 3 结果表明 HsCOBRAs 蛋白的二级结构均由无规则卷曲、α螺旋、延长链、及β折叠组成，所占比平均值分别为 46.02%、29.34%、18.95%、5.70%。6 个蛋白（HsCOBRA10、HsCOBRA19、HsCOBRA20、HsCOBRA21、HsCOBRA22、HsCOBRA23）中α螺旋占比最大（占 38.27%~58.25%），其次为无规则卷曲；其余 HsCOBRAs 蛋白中的无规则卷曲比例最大（占 38.67%~64.29%）；所有蛋白的β折叠所占的比例最小（占 1.55%~14.81%）。

表 3 蛇足石杉 COBRA 家族二级结构分析

Table 3 The secondary structure analysis of COBRA family of *Huperzia serrata*

蛋白 Protein	α螺旋 Alpha helix（%）	延长链 Extended strand（%）	β折叠 Beta turn（%）	无规则卷曲 Random coil（%）
HsCOBRA1	20.86	14.90	3.64	60.60
HsCOBRA2	14.44	23.35	4.67	57.54
HsCOBRA3	19.68	21.47	4.37	54.47
HsCOBRA4	14.44	23.35	4.67	57.54
HsCOBRA5	43.46	9.26	3.02	44.27
HsCOBRA6	31.15	18.03	1.64	49.18
HsCOBRA7	26.67	22.50	5.83	45.00
HsCOBRA8	32.39	14.79	7.75	45.07
HsCOBRA9	27.74	22.63	6.57	43.07
HsCOBRA10	38.27	20.99	14.81	25.93
HsCOBRA11	17.65	20.39	3.92	58.04
HsCOBRA12	19.48	13.64	2.60	64.29

HsCOBRA13	13.45	23.39	4.09	59.06
HsCOBRA14	15.99	22.62	4.98	56.41
HsCOBRA15	16.00	33.33	12.00	38.67
HsCOBRA16	26.25	22.50	7.50	43.75
HsCOBRA17	33.86	8.66	4.72	52.76
HsCOBRA18	25.29	21.26	2.30	51.15
HsCOBRA19	57.53	5.48	4.11	32.88
HsCOBRA20	43.97	17.02	7.80	31.21
HsCOBRA21	46.22	16.81	9.24	27.73
HsCOBRA22	48.71	10.34	7.33	33.62
HsCOBRA23	58.25	16.49	1.55	23.71
HsCOBRA24	12.31	31.54	7.69	48.46

2.5 蛇足石杉 *COBRA* 基因家族三级结构分析

利用 phyre^2 软件对蛇足石杉 COBRA 家族成员进行三级结构预测，表 4 和图 3 结果表明 24 个蛋白共有 19 种三级结构模型，HsCOBRA7 氨基酸比例最高（81%），HsCOBRA13、HsCOBRA16、HsCOBRA19、HsCOBRA24 均具有水解酶结构，HsCOBRA1、HsCOBRA6、HsCOBRA7 均具有水解酶抑制剂结构，HsCOBRA2、HsCOBRA3、HsCOBRA4、HsCOBRA14 均具有碳水化合物结构。HsCOBRA6、HsCOBRA7、HsCOBRA17 序列同源性为 99.9%；HsCOBRA6、HsCOBRA7 蛋白的三级结构相似，可能具有相似的生物学功能；HsCOBRA1、HsCOBRA5 序列同源性为 100%。

表 4 蛇足石杉 COBRA 家族三级结构分析

Table 4 The tertiary structure analysis of COBRA family of *Huperzia serrata*

蛋白 Protein	建模氨基酸及比例 Modeling amino acids and proportions (%)	模型 Model	序列同源性 Sequence homology (%)	模型名称 Model name
HsCOBRA1	135(45)	c2vt8B	100	水解酶抑制剂 Hydrolase inhibitor
HsCOBRA2	59（13）	c6f7eA	96	碳水化合物 Carbohydrate
HsCOBRA3	60（12）	c6f7eA	96.1	碳水化合物 Carbohydrate
HsCOBRA4	59（13）	c6f7eA	96	碳水化合物 Carbohydrate
HsCOBRA5	295(59)	c3vrhA	100	RNA 结合蛋白 RNA binding protein
HsCOBRA6	92(75)	c2vt8B	99.9	水解酶抑制剂 Hydrolase inhibitor
HsCOBRA7	97（81）	c2vt8B	99.9	水解酶抑制剂 Hydrolase inhibitor
HsCOBRA8	63(44)	c5zwo0	79.4	剪接 Splicing
HsCOBRA9	11(8)	d1v6ga2	36	糖皮质激素受体样结构域 Glucocorticoid receptor-like
HsCOBRA10	14(17)	c2kfvA	74.9	异构酶 Isomerase
HsCOBRA11	101(40)	d1exha	87.9	碳水化合物结合域 Carbohydrate-binding domain
HsCOBRA12	35(23)	c6qk4B	26.8	裂解酶 Lyase
HsCOBRA13	96(28)	c2rttA	90.4	水解酶 Hydrolase
HsCOBRA14	59(9)	c6f7eA	96.5	碳水化合物 Carbohydrate
HsCOBRA15	34(45)	d2a90a1	29.2	WWE 区域 WWE domain

HsCOBRA16	58(72)	c6uehA	68.5	水解酶 Hydrolase
HsCOBRA17	79(62)	c2l0bA	99.9	连接酶 Ligase
HsCOBRA18	28(16)	c2kk2A	34.3	信号蛋白 Signaling protein
HsCOBRA19	29(40)	c6i3oA	33.2	水解酶 Hydrolase
HsCOBRA20	15(11)	c3pmiC	77.3	结合蛋白 Protein binding
HsCOBRA21	22(18)	c6u9gA	22.8	转运蛋白 Transport protein
HsCOBRA22	15(6)	d1bboa1	18.2	β - β - α型锌指 Beta-beta-alpha zinc fingers
HsCOBRA23	22(11)	d2asba2	27	原核型 KH 结构域 Prokaryotic type KH domain
HsCOBRA24	41(32)	c3ndyG	66.2	水解酶 Hydrolase

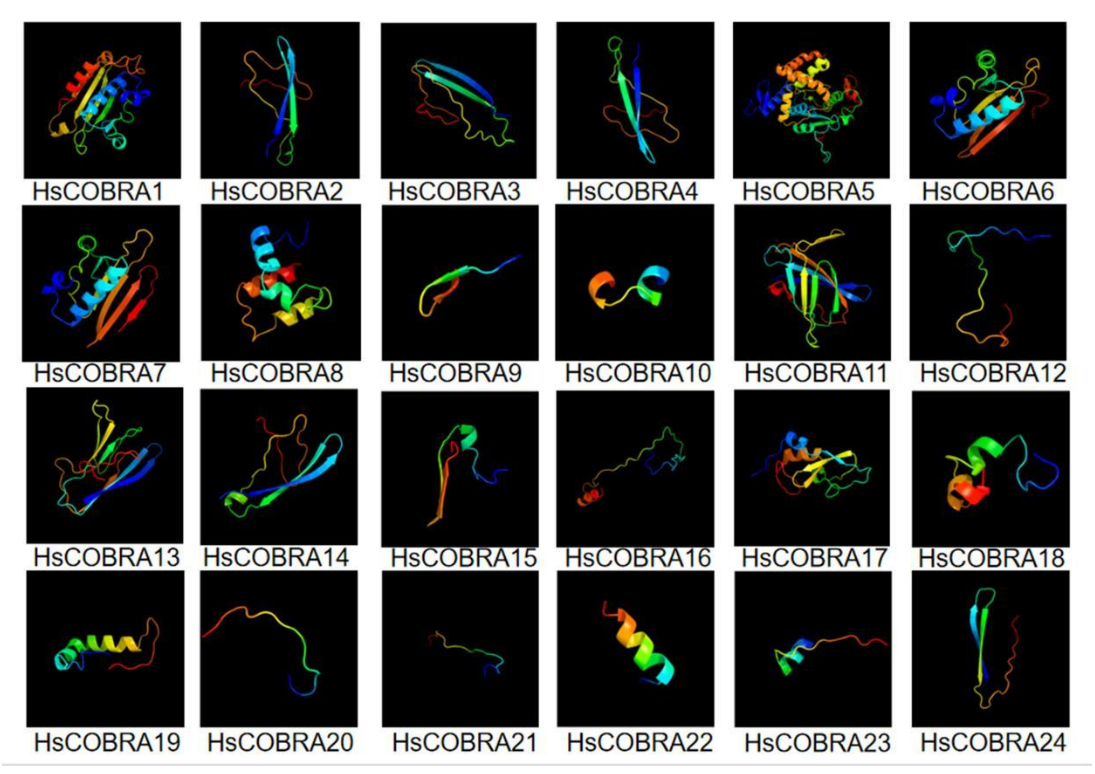


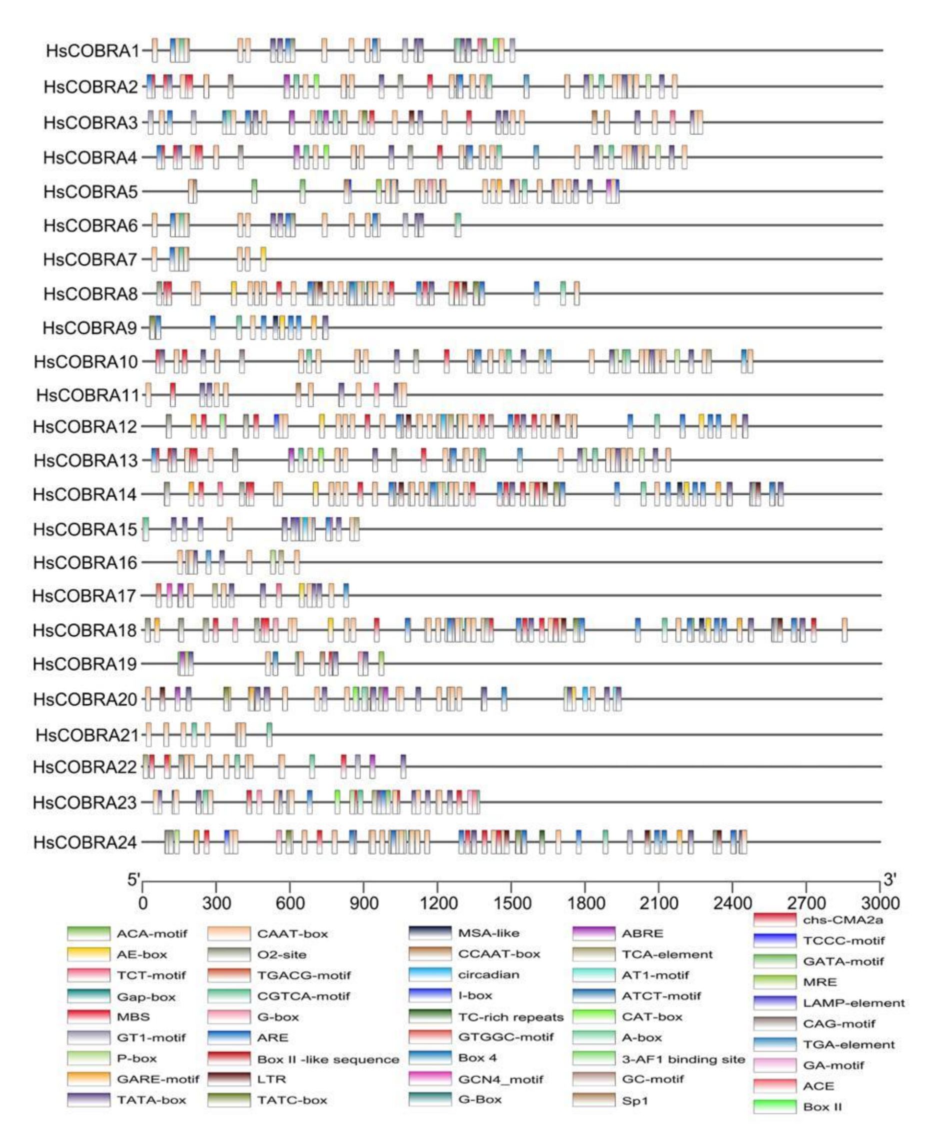
图 3 蛇足石杉 COBRA 家族三级结构预测图

Fig. 3 The tertiary structure prediction of COBRA family of *Huperzia serrata*

2.6 蛇足石杉 *COBRA* 基因家族顺式作用元件分析

顺式作用元件分析结果显示，HsCOBRAs 共有 46 种具有功能注释的顺式作用元件（图 4）。其中有 23 种顺式作用元件参与光响应（MRE、Box 4、G-box、ACE、circadian、AE-box、GT1-motif、Sp1、3-AF1 binding site、CAG-motif、GA-motif、Gap-box、TCT-motif、chs-CMA2a、chs-CMA2c、GATA-motif、LAMP-element、Box II、I-box、AT1-motif、ATCT-motif、GTGGC-motif、TCCC-motif）。部分蛋白具有参与分生组织表达（CAT-box）和胚乳表达（GCN4_motif）及响应脱落酸（ABRE）、生长素（TGA-element）、赤霉素（P-box、TATC-box）、水杨酸（TCA-element）的顺式作用元件，此外还含有响应厌氧（ARE）、缺氧（GC-motif）、干旱（MBS）、低温（LTR）等环境的顺式作用元件。CAAT-box 元件的数量最多（307 个），为启动子和增强子区域共有的顺式作用元件，分布于所有蛋白；其次是作为核心启动顺式作

用元件的 TATA-box（206 个）和参与干旱诱导 MYB 结合位点的 MBS（72 个）。HsCOBRA18 所含的顺式作用元件最多（78 个），其次为 HsCOBRA14（75 个）和 HsCOBRA24（70 个），HsCOBRA7 所含的顺式作用元件最少（10 个）。部分顺式作用元件仅在一条蛋白中显示，如 TCCC-motif（HsCOBRA5），3-AF1 binding site（HsCOBRA12），GTGGC-motif（HsCOBRA17），ACA-motif、Box II -like sequence（HsCOBRA18），Box II、chs-CMA2c（HsCOBRA19），ATCT-motif、AT1-motif、A-box（HsCOBRA20），GA-motif、ACE（HsCOBRA23），TC-rich repeats（HsCOBRA24）。



部分顺式作用元件重叠。

Partial *cis*-acting elements overlap.

图 4 蛇足石杉 COBRA 家族顺式作用元件预测图

Fig. 4 *Cis*-acting elements prediction of COBRA family of *Huperzia serrata*

2.7 蛇足石杉 *COBRA* 基因家族保守基序分析

HsCOBRAs 保守基序结果表明，该家族共包含 6 种保守基序，命名为 motif 1~motif 6（图 5，图 6）。其中，仅有 13 个蛋白拥有 motif，3 个蛋白（HsCOBRA2、HsCOBRA3、HsCOBRA4）所含的 motif 最多，均为 6 种；7 个蛋白（HsCOBRA1、HsCOBRA5、HsCOBRA7、

HsCOBRA9、HsCOBRA10、HsCOBRA12、HsCOBRA24）仅有 1 种 motif。motif 6 分布最广，其次为 motif 1、motif 3、motif 4。motif 3 含有高度保守的 CCVS 结构，7 个蛋白具有 motif 3。HsCOBRA1、HsCOBRA6、HsCOBRA7（motif 4），HsCOBRA2、HsCOBRA3、HsCOBRA4（motif 1~motif 6），HsCOBRA9、HsCOBRA10（motif 6）分别具有相同种类的 motif。

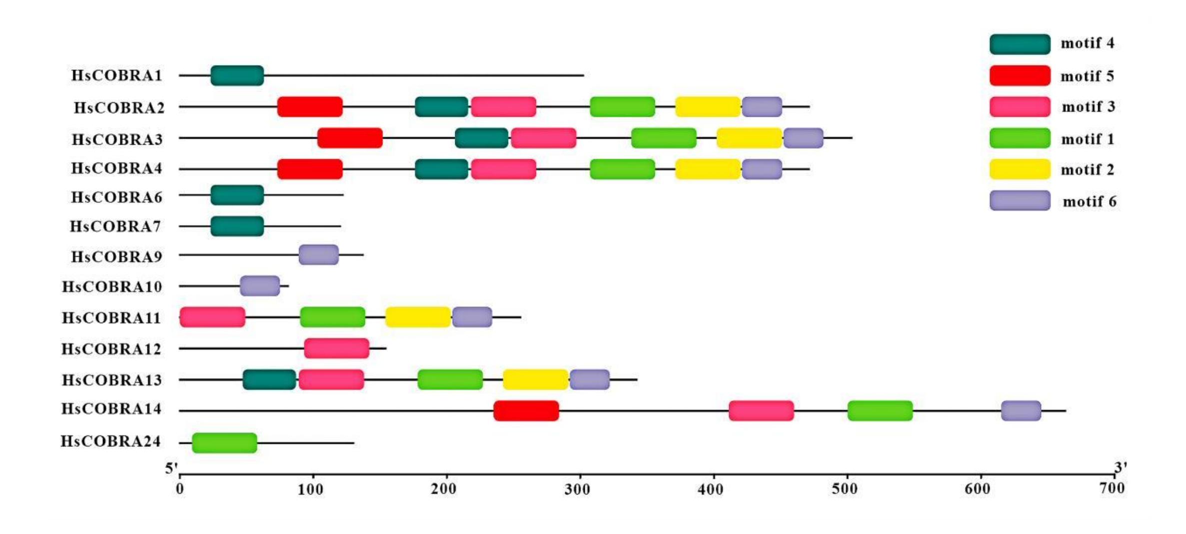


图 5 蛇足石杉 COBRA 家族保守基序预测图

Fig. 5 Conserved motif prediction of COBRA family of *Huperzia serrata*

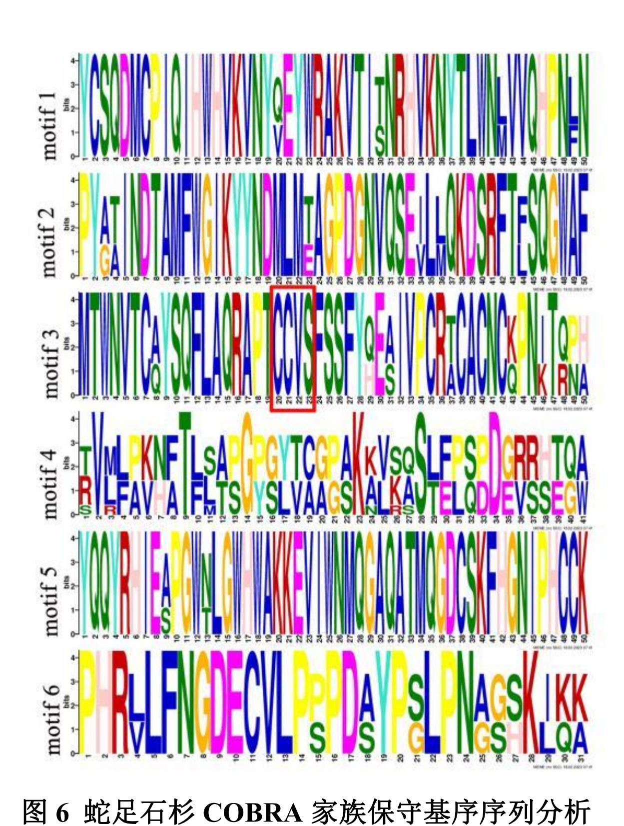
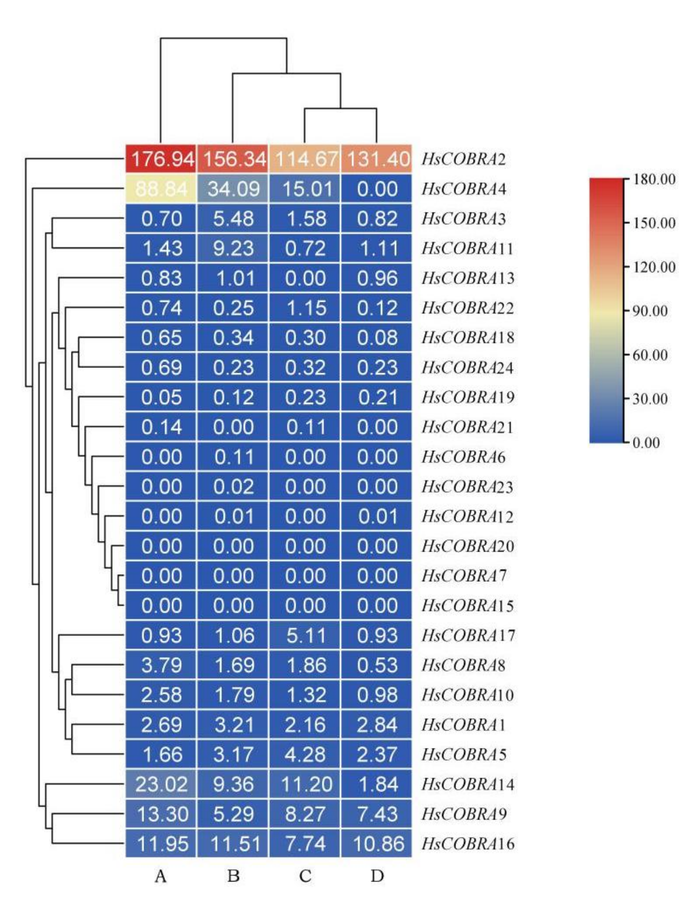


图 6 蛇足石杉 COBRA 家族保守基序序列分析

Fig. 6 Conserved motif sequence analysis of COBRA family of *Huperzia serrata*

2.8 蛇足石杉 COBRA 基因家族组织表达模式分析

根据 *COBRA* 基因在蛇足石杉叶、孢子、茎、芽胞中的转录组数据分析其表达模式，见图 7。*HsCOBRA*2 在 4 个组织中的表达量均最高，*HsCOBRA*4 在叶、孢子、茎中的表达量仅次于 *HsCOBRA*2。*HsCOBRA*7、*HsCOBRA*15、*HsCOBRA*20 在 4 个组织中几乎不表达，此外还有 3 个基因（*HsCOBRA*6、*HsCOBRA*12、*HsCOBRA*23）在叶片中几乎不表达，1 个基因（*HsCOBRA*21）在孢子中几乎不表达，4 个基因（*HsCOBRA*6、*HsCOBRA*12、*HsCOBRA*13、*HsCOBRA*23）在茎中几乎不表达，4 个基因（*HsCOBRA*4、*HsCOBRA*6、*HsCOBRA*21、*HsCOBRA*23）在芽胞中几乎不表达。*COBRA* 基因家族在蛇足石杉的中的表达量为叶＞孢子＞茎＞芽胞；*HsCOBRA*2 和 *HsCOBRA*16 在芽胞及孢子中表达量较高，可能参与蛇足石杉的繁殖。



A. 叶；**B**. 孢子；**C**. 茎；**D**. 芽胞。

A. Leaves; **B**. Spores; **C**. Stems; **D**. Gemma.

图 7 蛇足石杉 *COBRA* 家族的组织表达模式分析

Fig. 7 Expression profiles of *COBRA* family of *Huperzia serrata*

2.9 蛇足石杉 *COBRA* 基因家族系统发育分析

为了解蛇足石杉与蕨类植物 *COBRA* 基因家族的进化关系，构建蛇足石杉与江南卷柏（*Selaginella moellendorffii*）的 COBRA 家族成员的 NJ 系统发育树，见图 8。序列比对发现 HsCOBRA2、HsCOBRA3、HsCOBRA4、HsCOBRA11、HsCOBRA13、HsCOBRA14 具有 CCVS 结构，分别位于蛋白序列的第 238 位和第 241 位、第 268 位和第 271 位、第 238 位和第 241 位、第 20 位和第 23 位、第 109 位和第 112 位、第 431 位和第 434 位氨基酸残基中，蛋白高度保守。图 8 结果表明，HsCOBRA1 与 HsCOBRA7，HsCOBRA2 与 HsCOBRA4，

HsCOBRA16 与 HsCOBRA17 具有同源性；蛇足石杉与江南卷柏 *COBRA* 基因家族成员之间具有一定的亲缘关系。

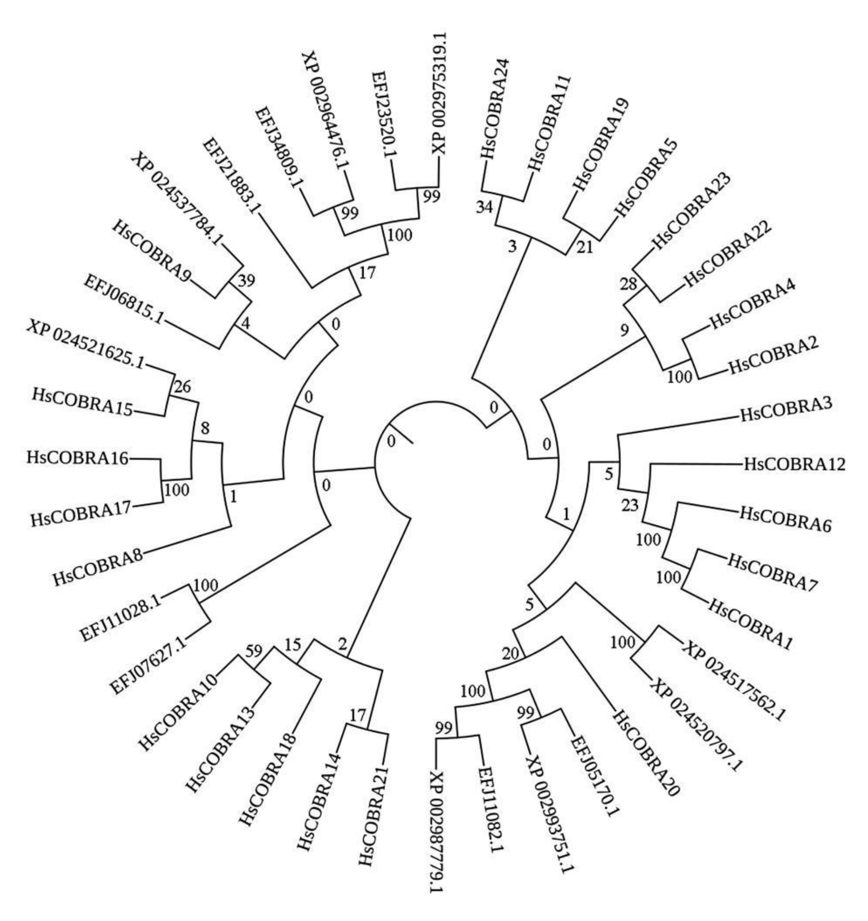


图 8 蛇足石杉与江南卷柏 COBRA 家族的系统发育树

Fig. 8 Phylogenetic tree of COBRA family of *Huperzia serrata* and *Selaginella moellendorffii*

3 讨论与结论

本研究共筛选出 24 个蛇足石杉 *COBRA* 基因家族成员，且多以碱性蛋白为主，二级结构主要由无规则卷曲和α螺旋及延长链构成，大部分成员具有光响应顺式作用元件，与大白菜(廉瑞婷等，2022）、亚麻(齐燕妮等，2019)、高粱（元志成等，2020）、番茄(曹颖等，2012)、玉米(潘倚天，黄敏，2022)、茶树(艾安涛等，2021)等 *COBRA* 基因家族相似，暗示这些物种中可能具有数量相近、性质和功能相似的家族成员。但研究也发现 HsCOBRAs 亚细胞定位与玉米、亚麻、谷子(任昂彦等，2021)中 COBRA 定位不同，提示不同物种的基因在进化过程中可能产生了某种特定功能的分化。HsCOBRAs 中 6 个蛋白（HsCOBRA15，HsCOBRA19，HsCOBRA20，HsCOBRA21，HsCOBRA22，HsCOBRA23）无结构域及保守基序，且在蛇足石杉四个组织中几乎不表达，该 6 个蛋白可能不属于蛇足石杉 COBRA 家族成员。

石杉碱甲在治疗阿尔茨海默病及益智中表现出时效长、副作用少的优势（陈思思等，2021），蛇足石杉作为石杉碱甲的主要来源之一，其原植物的生长繁殖、产量与可持续发展备受人们关注。*COBRA* 基因主要功能是参与植物细胞扩增(廉瑞婷等，2022），通过调节植物的生长发育从而影响其生物量和产量。研究中发现 HsCOBRAs 具有参与分生组织表达及胚乳表达的顺式作用元件，且 16 个 *HsCOBRAs* 在叶、孢子、茎、芽胞四个组织中有不同程

度的表达，尤其在叶及孢子中表达量最高，说明蛇足石杉 *COBRA* 基因家族参与到其叶、孢子、茎及芽胞的发育过程（杨兰锋等，2023）。前期研究发现，蛇足石杉大叶、老叶及孢子囊壁中均具有丰富的纤维素，提示 *HsCOBRAs* 基因家族可能通过调节纤维素及细胞壁生长、机械强度等参与蛇足石杉的生长发育过程，提高其生物量。孢子是蛇足石杉主要繁殖器官之一，*HsCOBRAs* 基因家族可能影响其孢子囊壁的生长发育从而影响着孢子的成熟和萌发。同时，该基因家族成员具有响应干旱、缺氧、低温、盐碱等非生物胁迫的顺式作用元件。其中 ACA 能响应冷、热、盐胁迫和紫外等多种非生物胁迫（陈逃飞等，2023）；ABRE 元件能与 BpbZIP1 特异性结合，响应盐胁迫，通过清除活性氧提高株系耐盐能力（郭依萍等，2020）；LTR 反转录转座子能响应盐、ABA、H_2O_2 和干旱等而发生转录水平的变化（周宾寒等，2023）。ACA、ABRE、LTR 反转录转座子在 *HsCOBRAs* 基因家族启动子区域大量存在，说明蛇足石杉 *COBRA* 基因家族在植物应对非生物胁迫方面发挥作用。

蛇足石杉 *COBRA* 基因家族具有与其他物种相似的生物信息学特征及规律，同时具有可能参与生长发育、繁殖及响应不良环境胁迫的生物学功能。通过研究蛇足石杉 *COBRA* 基因家族的分子生物信息学特征及组织表达规律，有助于阐明蛇足石杉 *COBRA* 基因的分子调控和促进生长发育机制，为生物量的增长提供新途径，为将来的分子育种及基因功能验证等研究提供理论依据。

参考文献：

艾安涛，李燕丽，章文益，等，2021. 茶树 COBRA 基因家族全基因组鉴定及表达分析[J]. 茶叶学报，62(3)：95-104. [AI AT, LI YL, ZHANG WY, et al., 2021. Genome-wide identification and expressions of COBRA family in *Camellia sinensis*[J]. Acta Tea Sin, 62(3): 95-104.]

曹颖，唐晓凤，刘永胜，等，2012. 番茄 COBRA 基因克隆、表达模式及生物信息学分析)[J]. 植物研究，32(3)：304-310. [CAO Y, TANG XF, LIU YS, et al., 2012. Cloning, expression pattern and bioinformation analyses of COBRA gene in Tomato(*Solanuml ycopersicum*)[J]. Bull Bot Res, 32(3): 304-310.]

陈思思，张梦华，王锦秀，等，2021. 药用植物千层塔的基原物种及研究进展[J]. 广西植物，41(11)：1794-1809. [CHEN SS, ZHANG MH, WANG JX, et al., 2021. Original plant and research progress of the medicinal plant *Huperzia javanica*[J]. Guihaia, 41(11):1794-1809.]

陈逃飞，黄敏，刘睿，等，2023. 玉米 ACA 基因家族生物信息学分析[J]. 分子植物育种，21(13):4188-4201. [CHEN TF, HUANG M, LIU R, et al., 2023. Bioinformatics analysis of maize ACA gene family[J]. Mol Plant Breed, 21(13):4188-4201.]

郭依萍，石晶静，周美琪，等，2020. 白桦 BpbZIP1 基因抗旱耐盐分析及 ABRE 元件结合鉴定[J]. 林业科学研究，33(5):68-76.[GUO YP, SHI JJ, ZHOU MQ, et al., 2020. Drought and salt tolerance analysis of BpbZIP1 gene in birch and abre element binding identification[J]. For Res, 33(5):68-76.]

黄玉妹，谢采连，周玲梅，等，2023. 千层塔基原及药效文献研究[J]. 亚太传统医药，19(3)：183-189. [HUANG YM, XIE CL, ZHOU LM, et al., 2023. Literature research on the original and efficacy of *Huperzia Serrata* (Thunb.) Trev. [J]. Asia-Pacific Trad Med, 19(3): 183-189.]

姬生国，2007. 国产石杉科药用植物的分子生物学研究[D].上海：复旦大学：1-140. [JI SG, 2007. Research on molecular biology of *Huperziaceae* medicinal plants[D]. Shanghai: Fudan University:1-140.]

廉瑞婷，任宏，翟文，等，2022. 大白菜 COBRA 基因家族鉴定及其低温胁迫表达分析[J]. 云南农业大学学报(自然科学)，37(6)：1031-1039. [LIAN RT, REN H, ZHAI W, et al., 2022. ldentification and expression analysis under low temperature stress of COBRA gene family in

Brassica rapa L. [J]. J Yunnan Agric Univ(Nat Sci), 37(6): 1031-1039.]

潘倚天，黄敏，2022. 玉米 COBRA 家族成员全基因组鉴定与表达模式分析[J]. 甘肃农业科技，53(2)：67-72. [PAN YT, HUANG M, 2022. Genome-wide identification and expression pattern analysis of COBRA family members in maize[J]. Gansu Agric Sci Technol, 53(2): 67-72.]

齐燕妮，李闻娟，王利民，等，2019. 亚麻 COBRA 基因家族的鉴定与生物信息学分析[J]. 甘肃农业科技，(9):33-38. [QI YN, LI WJ, WANG LM, et al., 2019. ldentification and bioinformatics analysis of the COBRA gene family in *Flax*[J]. Gansu Agric Sci Technol, (9):33-38.]

任昂彦，陈环宇，王根平，等，2021. 谷子 COBRA 家族成员的全基因组鉴定及表达模式分析[J/OL]. 分子植物育种:1-10[2023-08-05]. http://kns.cnki.net/kcms/detail/46.1068.S.2021 1015.0104.002.html. [REN AY, CHEN HY, WANG GP, et al., 2021. Genome-wide identification and expression files of COBRA family in foxtail mille[J/OL]. Mol Plant Breed:1-10[2023-08-05].http://kns.cnki.net/kcms/detail/46.1068.S.20211015.0104.002.html.]

王英全，梁景宏，贾瑞霞，等，2019. 2020-2050 年中国阿尔茨海默病患病情况预测研究)[J]. 阿尔茨海默病及相关病，2(1)：289-298. [WANG YQ, LIANG JH, JIA RX, et al., 2019. Alzheimer disease in China (2015- 2050) estimated using the 1% population sampling survey in 2015[J]. Chin J Alzheimer's Dis and Relat Disord, 2(1): 289-298.]

元志成，柯余生，吴富进，等，2020. 高粱 COBRA 基因家族全基因组的鉴定和表达分析[J]. 河南农业科学，49(10)：33-41. [YUAN ZC, KE YS, WU FJ, et al., 2020. Genome wide identification and expression of COBRA gene family in *Sorghum bicolor*[J]. J Henan Agric Sci, 49(10): 33-41.]

杨兰锋，朱旭东，周宾寒，等，2023. 苦荞 TCP 转录因子全基因组鉴定及非生物胁迫分析[J/OL]. 广西植物:1-15[2023-10-23]. http://kns.cnki.net/kcms/detail/ 45.1134.q. 2023 0824.1928.013. html. [YANG LF, ZHU XD, ZHOU BH, et al.,2023. Genome-wide identification and abiotic stress response analysis of TCP transcription factor gene family in tartary buckwheat (*Fagopyrum tataricum*)[J/OL]. Guihaia: 1-15[2023-10-23]. http://kns.cnki.net/kcms/detail/45.1134.q. 2023 0824.1928.013.html.]

周宾寒，杨竹，王书平，等，2023. 小麦幼苗活性 LTR 反转录转座子筛选及其对非生物胁迫的响应[J]. 作物学报，49(4): 966-977. [ZHOU BH, YANG Z, WANG SP, et al., 2023. Screening of active LTR retrotransposons in wheat(*Triticum aestivum* L.) seed lings and analysis of their responses to abiotic stresses[J]. Acta Agron Sin, 49(4): 966-977.]

BORNER GH, LILLEY KS, STEVENS TJ, et al., 2003. Identification of glycosylphosphatidylinositol-anchored proteins in *Arabidopsis*. A proteomic and genomic analysis[J]. Plant Physiol, 132(2):568-577.

JIA J, WEI C, CHEN S, et al., 2018. The cost of Alzheimer's disease in China and re-estimation of costs worldwide[J]. Alzheimerors Demen, 14(4):483-491.

KASIRAJAN L, HOANG NV, FURTADO A, et al., 2018. Transcriptome analysis highlights key differentially expressed genes involved in cellulose and lignin biosynthesis of sugarcane genotypes varying in fiber content[J]. Sci Rep, 8(1):11612.

LI P, LIU Y, TAN W, et al., 2019. Brittle culm 1 encodes a COBRA-like protein involved in secondary cell wall cellulose biosynthesis in *Sorghum*[J]. Plant Cell Physiol, 60(4):788-801.

LI Z, ZHOU T, SUN P, et al., 2022. COBL9 and COBL7 synergistically regulate root hair tip growth via controlling apical cellulose deposition[J]. Biochem Biophys Res Commun,

596:6-13.

NIU E, FANG S, SHANG X, et al., 2018. Ectopic expression of GhCOBL9A, a cotton glycosyl-phosphatidyl inositol-anchored protein encoding gene, promotes cell elongation, thickening and increased plant biomass in transgenic *Arabidopsis*[J]. Mol Genet Genomics, 293(5):1191-1204.

ROUDIER F, SCHINDELMAN G, DESALLE R, et al., 2002. The COBRA family of putative GPI-anchored proteins in *Arabidopsis*. A new fellowship in expansion[J]. Plant Physiol, 130(2): 538-48.

ZAHEER M, REHMAN SU, KHAN SH, et al., 2022. Characterization of new COBRA like (COBL) genes in wheat (*Triticum aestivum*) and their expression analysis under drought stress[J]. Mol Biol Rep, 49(2):1379-1387.